DOCUMENT RESUME

ED 435 871 CG 029 527

AUTHOR Rosnow, Ralph L.
TITLE Hedgehogs, Foxes, Ethics, and the Evolving Social Contract in Psychology.
PUB DATE 1999-08-00
NOTE 17p.; Paper presented at the Annual Convention of the American Psychological Association (107th, Boston, MA, August 20-24, 1999).
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ethics; Higher Education; *Psychological Studies; Psychology

ABSTRACT

This paper discusses the importance of ethics in psychological research. It defines the social contrast between psychological science and society as the responsibility not to do psychological or physical harm to any research participants and to do beneficial research in a way that will produce valid research. Also explored are ways in which ethical issues encountered during research activities can be handled. This paper examines the idea that squandering scientific opportunities in poorly designed and weakly analyzed research makes for bad ethics because scientific data are expensive in terms of time, effort, money, and other resources. This waste management concept is examined within the framework of four metaphorical principles. (Contains 20 references.) (MKA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 871

Hedgehogs, Foxes, Ethics, and the Evolving Social Contract in Psychology

George A. Miller Award Presentation

APA Convention, Boston, Aug. 21, 1999

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Ralph L. Rosnow

Department of Psychology

Temple University

Philadelphia, PA 19122

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. Rosnow
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

I am deeply grateful to Division 1 for honoring me with the George A. Miller Award. The article designated for this award was published in *Psychological Methods* in 1997. I thank Mark Appelbaum, editor of the journal, who invited me to submit the paper, and Howard Sandler, who patiently shepherded it through the review process. I also thank Temple University for the continuous support that I have received through the Thaddeus Bolton Professorship.

Emerson once wrote: "We want but two or three friends, and these we cannot do without, and they serve us in every thought we think." Mimi Rosnow, Bob Lana, and Bob Rosenthal are for me the embodiment of that sentiment. Bob Rosenthal and I first met when we were young assistant professors at Harvard University and Boston University, respectively. Over 35 years, Bob has improved my thinking about research methods to fill a baker's dozen books together, including our most recent, coauthored with Don Rubin, *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach*. Later on, I will have a little more to say about this approach. Bob Lana, my Ph.D. adviser at American University from 1960-1962, and my Temple University colleague for 32 years, whetted my interest in the epistemological foundations of psychology, introduced me to the joys and complexities of experimental social psychology, and is the reason I moved to Temple University in 1967, to help him develop a new doctoral program in

CG029527

BEST COPY AVAILABLE

social psychology. Mimi Rosnow has been my constant intellectual and emotional companion longer than I am permitted to tell you. She is the creative person who conceived the terms "wish rumor," "dread rumor," and "synthetic benevolence," and is the reason why many students find *Writing Papers in Psychology* so accessible. I am profoundly indebted to all three for their generosity of spirit, penetrating insights, and tireless mentoring, which have enriched my life and my work.

Some years ago, George A. Miller wrote that the role of scientific psychology was to promote human welfare, and urged individual psychologists to instill scientific facts about the potential of human nature in the public consciousness. My article in *Psychological Methods* was addressed in large part to new researchers who, in their efforts to use the special tools of science to produce such facts, felt burdened by an expanding body of bureaucratic rules and regulations. Of course, even old researchers can find themselves caught in a bind between the methodological demands of science and the moral sensitivities of society. The hedgehogs-and-foxes reference in the title goes back to the poet Archilocus, who in 650 B.C. stated, "The fox knows many things, but the hedgehog knows one big thing." I first encountered this reference in a classic essay by Isaiah Berlin, who compared Dante, Plato, Hegel, and Nietzche's single central focus with Shakespeare, Aristotle, and Goethe's pursuit of many visions on many different levels. My generation of researchers (and earlier ones) belonged to the hedgehogs, embracing a single central vision of science as an "endless frontier," to quote Vannevar Bush's famous phrase. Because researchers today must navigate labyrinthine byways of an evolving social contract of moralistic do's and don'ts, and still comply with technical scientific demands, they must become foxes, that is, move on many levels and pursue many visions.

Broadly speaking, the social contract between psychological science and society can be described as the responsibility not to do psychological or physical harm to any of our research participants and to do beneficent research in a way that will produce valid conclusions. Often, however, this is easier said than done. Confidence in the validity of scientific facts may be

jeopardized by the suspicion of subject or experimenter artifacts, but controlling for artifacts calls for a delicate balance between methodological and ethical concerns.

Recently, Bob Rosenthal and I revisited this theme in *People Studying People: Artifacts and Ethics in Behavioral Research*. We began by quoting the social scientist Otto Neurath, one of the founders of the Vienna Circle:

> We are as sailors who are forced to rebuild their ship on the open sea, without ever being able to start fresh from the bottom up. Wherever a beam is taken away, immediately a new one must take its place, and while this is done, the rest of the ship is used as support. In this way, the ship may be completely rebuilt like new with the help of the old beams and driftwood--but only through gradual rebuilding.

We added that, just as old beams must constantly be replaced by new ones, so must the old sailors be replaced by the new. The new sailors must learn from the old what factors weaken the beams and how they are to be replaced. Because artifacts in research weaken its beams, each generation of sailors must learn to spot them in order to improve the ship of science. Although replacing old beams is essential, it can be dangerous. Passengers can be hurt as the ship is improved, so sailors must learn how to sail safely, harming no one aboard. Understanding the capabilities and limitations of our methodological tools helps keep the ship seaworthy; sensitivity to moral values keeps the sailors worthy and the passengers safe.

That there are moral underpinnings of science is now generally taken for granted. This was not always the case, especially during the heyday of the positivist movement. According to the prevailing image, scientists were disinterested seekers after truth, characterized by a "see no evil, hear no evil" impartiality. Science is morally neutral by its very nature, it was said, because the moment it begins sorting facts into "good ones" and "bad ones" it is no longer science. Some theorists drew a distinction between natural and behavioral science, arguing that although moral detachment was conceivable in natural science, it was impossible in behavioral science because of the societal implications of psychological facts. For example, when we speak of behavior as

"normative," the implication to a layperson is that we mean the behavior is to be expected, or is morally desirable. When we disseminate facts about prejudice, mental illness, or child abuse, we are touching on societal issues that are highly charged with moralistic implications. Even when we comment on topics using facts that appear to us to be morally neutral (learning behavior, for example), to others those facts may be supercharged with values and conflicts.

As the history of science teaches, the assumption that natural science is morally neutral crumbled with the development of atomic physics. Ultimately, unrest surfaced in all disciplines about the motives, purposes, and moral implications of research itself. In psychological science, going back to the 1960s, leading spokespersons had voiced concerns about the status of human values in research. Like the impassioned plea of the Tin Woodman in *The Wizard of Oz*, who prayed to be given a heart, there were ardent pleas to give our science a humanistic heart. In 1966, the APA created a task force which was assigned to compose a code of ethical principles for research with human participants. Out of those deliberations came the ten basic guidelines which researchers have euphemistically referred to as the "Ten Commandments of APA." A favorite quote from that period was Kenneth Gergen's forewarning:

> Most of us have encountered studies that arouse moral indignation. We do not wish to see such research carried out in the profession. However, the important question is whether the principles we establish to prevent these few experiments from being conducted may not obviate the vast majority of contemporary research. We may be mounting a very dangerous cannon to shoot a mouse.

What Ken labeled a "dangerous cannon" now seems like a popgun in light of pressures, obstacles, and priorities imposed on researchers by daunting arrays of ethical considerations, bureaucracies, formalities, and legalities that did not exist in their present form in the 1960s. As the APA's proposed revision of its ethical guidelines for research implies, virtually every aspect of the scientific agenda may be perceived as value-laden to some degree, from the selection of a

research topic, through the conceptualization and implementation of the study, to the statistical analysis, interpretation, and reporting of results.

In his recent book, *Ethical Issues in Behavioral Research*, Allan Kimmel discussed the various codes of ethics that have been adopted by psychological associations in the U.S., Canada, France, Germany, Great Britain, the Netherlands, Poland, and elsewhere. Clearly, psychological research around the globe is bounded by professional rules and guidelines. In the U.S. we have become accustomed to having IRBs play the role of gatekeepers, supervising the flow of institutional research. A few years ago, while serving on the APA's Committee on Standards in Research, Mary Jane Rotheram-Borus, Steve Ceci, Peter Blanck, Gerry Koocher, and I wrote in the *American Psychologist* about the vagaries and limitations of the decision process involved in IRB ethical evaluations. The idealized decision-plane model in the figure, which Bob Rosenthal and I first published some years ago, will help me explain one aspect of this problem. It shows how the cost of doing research might be simultaneously evaluated against the utility of doing it. Presumably, a study falling in the region labeled "D" (that is, high utility and low cost) would be approved by an IRB, whereas a study in the area labeled "A" (low utility and high cost) would be rejected. On the B-C axis would be studies in which the perception is that the costs and utilities are balanced, thereby changing the decision from easy to difficult. In the case of low-cost, low-utility research, however, an IRB might be unwilling to approve a study that it viewed as harmless but unlikely to yield any substantial benefit.

As most of us know from personal experience, the gatekeeper process is far less reliable. Some years ago, Steve Ceci, Douglas Peters, and Jonathan Plotkin sent sample research proposals to the chairpersons of several hundred IRBs, asking for their candid assessments. The proposals outlined a research investigation of discrimination in the hiring of managerial positions by Fortune 500 corporations. Some proposals, by mentioning instances of discrimination or reverse discrimination, were intended to appear highly sensitive; others, by documenting discrimination based on height or weight, were intended to appear less sensitive. The proposals constructed to

seem highly sensitive were twice as likely to be rejected by the IRBs, the primary reason given for rejection being the political impact of the anticipated results. It frequently seems that IRBs ignore utilities and merely use the A-C axis value as their criterion, as in this case. But even when utility is a consideration, the idealized model is insufficient because it ignores the costs of research not done. By concentrating only on the act of "doing research," and ignoring the act of "not doing research," the decision process concedes to a less rigorous standard of accountability than that aspired to by most researchers. Another example would be if an investigation designed to reduce the risk of HIV infection were rejected on the grounds that the proposed methodology could not guarantee the privacy of the participants. Likewise, rejecting a study that involved a deception experiment to reduce violence or prejudice would not solve the ethical problem, but would merely trade one moral issue for another.

I do not mean to imply that researchers should try to circumvent the review process, but (as discussed more fully in my article) I think we need to generate ideas and implement a method to improve this process. The Committee on Standards in Research was dissolved by the APA in 1993, but in our final report we made an appeal for the development of a mechanism to keep IRBs abreast of emerging issues in psychological science and to raise their consciousness about the costs to science and society of sensitive research not done. We also argued for speedy access to appeals when IRBs are subject to political pressures that restrict their judgments.

As Donald Bersoff, Joan Sieber, and other leading psychologists have observed, there are exploitable research opportunities in ethical dilemmas. Twenty-five years ago, Bob Rosenthal and I alluded to one such opportunity in our book *The Volunteer Subject*. In an early example of meta-analysis (before the term was coined by Gene V Glass), we used pooled empirical data to tease out ethical-enhancing strategies for reducing volunteer bias by stimulating research participation. Daniel E. Koshland, in an editorial in *Science* magazine a few years ago, noted a medical case in which, it would seem, such strategies could be advantageous. Suppose a study of cholesterol in the diet needed volunteer subjects, but because people at high risk were most likely

to volunteer, it could be dangerously risky to generalize from this group to a more normal one. On the other hand, if there were ethically acceptable ways to improve the rate of participation by the reluctant nonvolunteers, that should also improve the generalizability of the research sample. In our book, Bob and I recommended a number of such strategies, including making the appeal as interesting as possible to capture people's attention, making it as nonthreatening as possible to avoid eliciting unwarranted fears, and emphasizing the theoretical and practical significance of the research to encourage willingness to participate. Using these empirically-justified strategies should not only reduce the amount of volunteer bias, but also make us more thoughtful scientists, because they will improve our relations with participants. Moreover, if we were to tell potential participants as much as possible about the significance of our research--as though they were another granting agency, which in fact they are, granting us time instead of money--we would have to do significant research.

In the remainder of my talk I would like to examine another facet of the ethical dilemma, which I call the "waste not, want not problem," based on an argument that Bob Rosenthal raised in *Psychological Science* a few years ago. He argued that squandering scientific opportunities in poorly designed and weakly analyzed research makes for bad ethics because scientific data are expensive in terms of time, effort, money, and other resources, and he mentioned a number of ways to improve this situation. Consistent with that waste management objective, I will examine four aspects of this problem within the framework of four metaphorical principles: the reader's lament, the scientist's key, Tarzan's leap, and the dayyan's decree.

First, one way that waste can occur is when investigators succumb to what statisticians James O. Berger and Donald A. Berry called "the illusion of objectivity" in statistical analysis, unwittingly work with low power, and conclude there was "no effect" merely because they found p to be on the "unpopular" side of .05. Perhaps we need to remind ourselves and our students of the following conceptual relationship: significance test = size of effect × size of study. It shows

that significance tests (such as *t*, *F*, and chi-square) can be understood as the product of two components, one having to do with the size of the effect (as measured, for example, by Cohen's *d*, Hedges's g, or the product-moment *r* between the independent variable and the dependent variable) and the other with the size of the study (i.e., the number of sampling units). The equation reminds us that it is easier to claim the presence of a phenomenon of any given magnitude at the desired significance level when working with a larger *N* than with a smaller *N*. Thus it is crucial to realize that finding "nonsignificance" is not the same thing as finding "no effect," as it is quite possible for a meaningful effect to be present although the significance test lacks sufficient power to detect it at the desired significance level due to modest *N* or an imprecise design. By the reader's lament, I mean that significance testing with insufficient power is like switching off the light just as you sit down to read a good book. Claiming there is no book would be deceptive, and wasteful. Reporting the magnitude of any obtained effect (and its confidence interval or null-countermull interval), or providing the raw ingredients so others can compute these values, allows research consumers to make up their own minds about whether there was any meaningful effect.

Second, another problematic situation concerns the bias against publishing replications, which is further exacerbated by the publish or perish tradition because it encourages researchers to submit single studies. It is a giant leap of faith, however, to assume that because one found a statistically significant result, it means the result is reproducible and generalizable. As practical experience has taught most of us, a psychology experiment conducted with one group of subjects in one place at one time may yield results very different from the "same" experiment conducted with another group of subjects in another place at another time. Even though reproducibility is almost universally accepted as the most important criterion of genuine scientific knowledge, it is precariously rare in some areas of our science. In a study reported in the *International Journal of Research in Marketing* in 1994, Raymond Hubbard and J. Scott Armstrong noted that of 1120 papers they sampled from three influential journals in marketing research, none were replications.

Just under 2% were what they called "extensions," of which only 3 papers provided full confirmation of the original results. Without repeating experiments we cannot, as Bill McGuire wisely recommended, continue the discovery process by clarifying and expanding the meaning and limits of hypotheses. Replication can be understood as a waste control strategy because, as Hubbard and Armstrong observed, it protects the literature from the uncritical acceptance and dissemination of erroneous and questionable results. We have a collective responsibility, they concluded, "to ask whether a given result is plausible, reproducible and/or generalizable."

This brings me to the principle I called "the scientist's key," which takes its name from something that James B. Conant, an experimental scientist, once wrote. He said the scientist is like someone who is trying to unlock a door using a hitherto untried key. The role of replication, we might say, is to make the key available to competent others so they can learn for themselves what is behind the door. I do not mean reproducing the identical *p* value, but observing a similar relationship or seeing the same phenomenon. Suppose you were jogging around Boston Common one morning and spotted two Martians--not two people disguised as Martians, but real Martians: green skin, antennas poking out of their scalps, etc. You aren't going to whip out your calculator, but you sure are going to ask somebody, "Do you see what I see?"

Third, it would also be prudent to temper our overreliance on omnibus statistical tests, which we were taught as graduate students to regard as protection from the dangers of "data mining." Unfortunately, omnibus tests can trick unsuspecting researchers into abandoning the focused question of interest. To illustrate, suppose a student who was working on his thesis at Seattle University was interested in assessing two theories, A and B, each implying a specific prediction about how many counseling sessions it will take to improve psychological functioning in parents of children with serious illness. Theory A predicts a minimum of four sessions to produce any benefit, and that any fewer than four will be futile. Theory B predicts small benefits as early as the first session, with gradual improvement expected to continue throughout all four sessions. The student designs a randomized experiment consisting of four groups, corresponding

to 1, 2, 3, or 4 sessions of counseling. So far, so good. Unfortunately, instead of inspecting his results more closely, the student computes an omnibus *F* test. To his dismay, he finds the associated *p* is not significant at .05, assumes it means there was no effect, and ends up sleepless in Seattle worrying that he will never get a degree.

The trouble with omnibus statistical tests is that they seldom tell us anything we really want to know. All the while that a particular predicted pattern may have been evident to the naked eye, the student's reliance on an omnibus *F* may have led him to miss the forest for the *p*. Instead of addressing the two focused questions of interest, he addressed a diffuse question of dubious scientific value. Science is not a Simon says game in which researchers must seek permission from the *p* value associated with an omnibus *F* before they can look for answers to specific questions. The student's 3 *df-F* was too nebulous to be of value to him, because the omnibus *F* would be the same whether he were interested in the prediction implied by Theory A or Theory B. Moreover, informative effect size indices such as Cohen's *d*, Hedges's *g*, and the product-moment *r* cannot be gotten from an omnibus *F*. Relying on omnibus *F* tests can be hazardous for many researchers, and may become even more perilous when there are several dependent variables and multiple degrees of freedom for the independent variables. This leads me to the principle I call Tarzan's leap, which refers to something Johnny Weissmuller, who played Tarzan in the movies, said. Asked about the nature of the apelike skills he displayed while swinging through the trees, he replied "the main thing is not to let go of the vine." It is good advice for researchers who let go of their specific hypotheses prematurely, often without ever testing them.

Returning to the insomniac student, what he should have done was to calculate contrasts, which would have allowed him to address the competing predictions in a precise way. To do so, he would begin by expressing the predictions as integer lambda values that sum to zero. Theory A predicted no benefits prior to four sessions, but a substantial benefit after Session 4, which can be stated in terms of lambda weights of -1, -1, -1, +3. Theory B predicted a continuous linear

increase of benefits, which can be expressed by lambda weights of -3, -1, +1, +3. Now all that is needed is to plug the weights into the correct formulas, calculate meaningful indices of effect size, construct confidence intervals or null-counternull intervals around those effect size indices, and interpret the results. If you are not familiar with the counternull value, it refers to the non-null magnitude of the effect size that is supported by the same amount of evidence as is the null value of the effect size. The null-counternull interval implies whether conclusions of "no effect" might be in error, thereby providing some protection against the anguish that may be experienced by those who balance their "accept/reject" decisions on the razor's edge of the 5% *p*.

Incidentally, an easy way to create appropriate λ weights is to estimate the mean outcome in each group, then subtract the overall mean from each group mean to produce weights that sum to zero. In the case of Theory A, suppose we predicted condition means of 0, 0, 0, 4 for sessions 1, 2, 3, 4, respectively. Subtracting the overall mean of 1 from the four means gives us weights of -1, -1, -1, +3. For Theory B, suppose we had predicted condition means of 1, 2, 3, 4 for the four "dosage" levels. Subtracting the overall mean of 2.5 gives us -1.5, -0.5, +0.5, +1.5, which multiplied by 2 gives us lambda weights of -3, -1, +1, +3. A useful post hoc procedure (just as easy to do) is to correlate the weights and the obtained group means. Bob Rosenthal, Don Rubin, and I call this aggregate correlation an "alerting *r*" because it can alert us to overall trends of interest. Another useful feature of the alerting *r* is that squaring it reveals the proportion of the overall between-condition sum of squares that can be accounted for by the particular set of weights. With this information, we can easily carve a contrast out of the omnibus *F* to address the predicted pattern. We can find out whether, given the particular circumstances of the investigation, Theory A was a better predictor than B, or B was better than A, or neither theory did particularly well, or both theories did equally well.

Finally, another cost conscious, waste management strategy would be to allow for the distinct possibility of more than one valid theoretical explanation. You may be familiar with the

scientific approach called "strong inference" by the physicist John R. Platt, in which the strategy is to pit one theory against another in a game of empirical jeopardy and eliminate the loser. This approach may have worked in the gladiatorial combat of ancient Rome, and may work well in physics, but it can be wasteful in psychological science because behavior is often "pushed" and "pulled" by more than one determinant. The principle of the dayyan's decree is sensitive to this problem. The name comes from a Yiddish anecdote about a rabbinical judge who was asked by a married couple to settle a conflict in which they were embroiled. The wife told her side and the dayyan said, "You are right." The husband told his side and the dayyan again said, "You are right." An incredulous talmudic student who overheard the conversation addressed the dayyan: "Rebbe, you really mean they are both right?" The dayyan replied, "You are right, too." Many years ago, Donald Campbell and Julian Stanley advised researchers that when two well-grounded theories disagree, both may be right to some degree. The lesson of the dayyan's decree is not to assume there is only one valid explanation, and lambda weights and contrasts can again serve us exceedingly well by helping us ascertain the degree to which each theory is applicable in a given situation. Likewise, if we were interested in finding out how well the two competing predictions fare together, we could compute a contrast using the combined, *Z*-scored lambdas.

I began by referring to one facet of the work of George A. Miller, after whom this award is named, and it is fitting that I conclude by mentioning another side of his work. Thirty years ago, writing in the *American Psychologist*, he proposed a code of priorities for assessing the costs and benefits of new technologies before they were introduced to society at large. Another giant of psychology, Gordon Allport, estimated the life of popular concepts in psychology to be about two decades, after which, he said, "they begin to taste as flat as yesterday's beer." Miller's work was far above average; the eleven criteria he proposed in that article remain as vital today as they were three decades ago: validity, intelligibility, reliability, social relevance, safety, accountability, the use of informed consent, avoidance of deception, and emphasis on individuality, availability, and distributability. I would suggest adding a twelfth to that list, precision in specification, so

that both we and society have an accurate picture of what we are prepared to instill in the public consciousness. Contrasts and effect sizes encourage us to be precise in our thinking and writing, and provide an easy-to-use methodology for doing so. Used wisely and ethically, they will allow us to go a long way in fulfilling the benevolent vision that Miller articulated so eloquently.

References

Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist, 76,* 159-165.

Berlin, I. (1953). *The hedgehog and the fox: An essay on Tolstoy's view of history.* NY: Simon & Schuster.

Ceci, S. J., Peters, D., & Plotkin, J. (1985). Human subjects review, personal values, and the regulation of social science research. *American Psychologist, 40,* 994-1002.

Gergen, K. J. (1973). Codification of research ethics: Views of a doubting Thomas. *American Psychologist, 28,* 907-912.

Hubbard, R., & Armstrong, J. S. (1994). Replications and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing, 11,* 233-248.

Kimmel, A. J. (1996). *Ethical issues in behavioral research: A survey.* Cambridge, MA: Blackwell.

Koshland, D. E., Jr. (1993). Science and society. *Science, 260,* 143.

McGuire, W. J. (1986). A perspectivist looks at contextualism and the future of behavioral science. In R. L. Rosnow & M. Georgoudi (Eds.), *Contextualism and understanding in behavioral science: Implications for research and theory* (pp. 271-301). NY: Praeger.

Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist, 24,* 1063-1075.

Miller, G. A. (1970). Assessment of psychotechnology. *American Psychologist, 25,* 991-1001.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347-353.

Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science, 5,* 127-134.

Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject.* NY: Wiley.

Rosenthal, R., & Rosnow, R. L. (1984). Applying Hamlet's question to the ethical conduct of research: A conceptual addendum. *American Psychologist, 39,* 561-563.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* NY: Cambridge University Press.

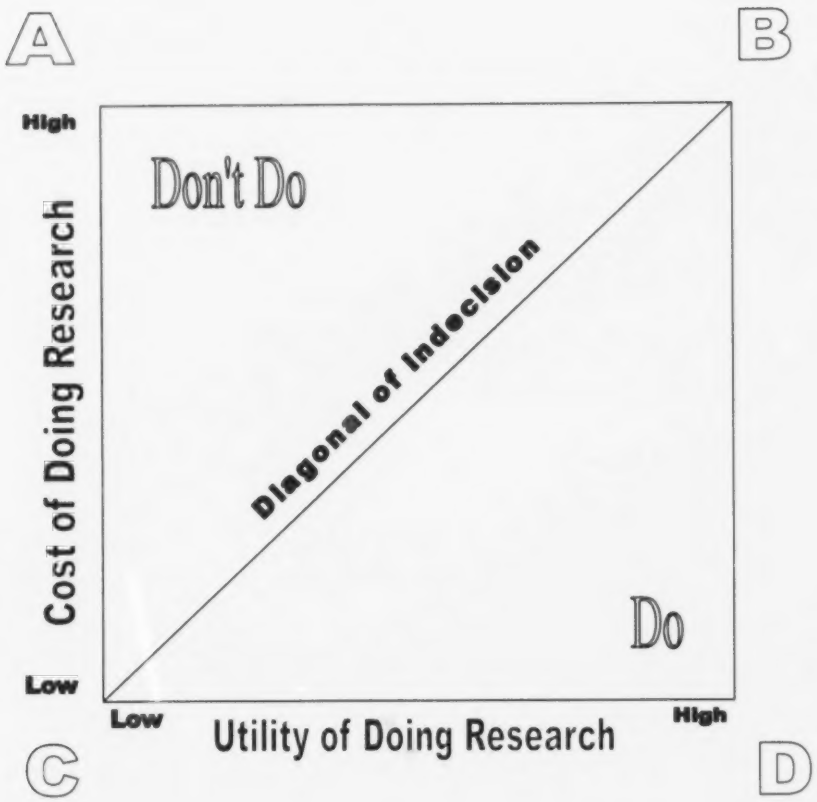
Rosnow, R. L. (1997). Hedgehogs, foxes, and the evolving social contract in psychological science: Ethical challenges and methodological opportunities. *Psychological Methods, 2,* 345-356.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1,* 331-340.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research.* NY: Freeman.

Rosnow, R. L., & Rosnow, M. (1998). *Writing papers in psychology: A student guide* (4th ed.). Pacific Grove, CA: Brooks/Cole.

Rosnow, R. L., Rotheram-Borus, M. J., Ceci, S. J., Blanck, P. D., & Koocher, G. P. (1993). The institutional review board as a mirror of scientific and ethical standards. *American Psychologist, 48,* 821-826.



BEST COPY AVAILABLE